Course Name: Certificate Course in Big Data & Machine Learning with Python Using Azure

Course Objective: This course provides a comprehensive and in-depth introduction to the Big Data and Machine Learning. It covers a global perspective, explores real-world examples, and delves into future trends. With case studies and hands-on projects, participants will gain practical experience and valuable insights to apply in real-world scenarios.

Pre – Requisite: No prior programming skills are required. Access to certain combinations of hardware is required

Course Outcome: Students will learn to build, train, and deploy machine-learning models to solve real-world problems, gaining expertise in algorithms, data analysis, and predictive modelling.

Course Duration: 80 Hrs (8 hours/ day for 2 Weeks)

Teaching Schema:

S. No.	Modules	Hours
1	Introduction to Data Science and Machine Learning	4
2	Exploratory Data Analysis (EDA) using Numpy & Pandas	8
3	Data Visualization Libraries, Data Engineering and Pre-processing	10
4	Cloud Computing for ML & Environment Setup with Python and Azure	6
5	Machine Learning & Model Development with Azure ML and Visual Studio Code	10
6	Machine Learning Algorithm	12
7	Model Evaluation and Validation Strategies & Handling Overfitting and Underfitting	6
8	Model Parameter and Hyperparameter Tuning	8
9	Prescriptive Analytics and Optimization	8
10	Capstone Project	8
	Total	**80**

Detailed Course Content:

1: Introduction to Data Science and Machine Learning

- Definition and Scope of Data Science & Machine Learning
- Applications in Artificial Intelligence and Machine Learning
- Supervised, Unsupervised, and Reinforcement Learning
- Overview of Basic Regression and Classification Algorithms

2: Exploratory Data Analysis (EDA) using Numpy & Pandas

- Introduction to Pandas and Numpy
- Explanation of Key Data Structures: Series and DataFrame
- Hands-on Exploration: Summarizing, Filtering, and Transforming Data
- Data Cleaning Techniques: Handling Missing Values and Outliers
- Statistical Analysis of Data Using Numpy
- Slicing/Dicing Operations

3: Introduction to Data Visualization Libraries

- Introduction to Matplotlib & Seaborn
- Creating Graphs: Bar Charts, Pie Charts, Line Charts, Histograms, Scatter Plots
- Advanced Plotting Techniques: Heatmaps, Pair Plots, Categorical Plots

4: Data Engineering and Pre-processing

- Data Cleaning and Processing
- Feature Selection & Feature Engineering
- Handling Outliers and Missing Data
- Cross-Validation Strategies
- Data Scaling and Normalization: Standardization, Min-Max Scaling

- Dealing with Categorical Variables: One-Hot Encoding, Label Encoding

5: Fundamentals of Cloud Computing for ML

- Integration of Cloud Computing with Machine Learning
- Overview of AWS, Azure, and GCP ML Services

6: Setting Up the Environment with Python and Azure

- Setting up a Python Development Environment
- Use of the Azure SDK for Python
- Integration with Visual Studio Code

7: Machine Learning Fundamentals & Model Development

- Introduction to Supervised & Unsupervised Learning
- Machine Learning Problem Statement and Analysis
- Regression vs Classification: Use Cases and Examples
- Understanding Overfitting and Under fitting
- Model Deployment using the Azure ML SDK and Visual Studio Code

8: Supervised Learning – Regression & Classification

Regression

- Introduction to Regression: Definition, Types, and Use Cases
- Linear Regression: Theory, Cost Function, Gradient Descent
- Evaluation Metrics: MSE, R-Squared, MAE

Classification

- Introduction to Classification: Types and Use Cases
- Logistic Regression: Theory, Logistic Function, Binary & Multiclass Classification
- Decision Trees: Construction, Splitting Criteria, Pruning, and Visualization
- Random Forests: Ensemble Learning, Bagging, and Feature Importance
- Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, ROC Curves
- Implementation of Classification Models using Scikit-Learn

9: K-Nearest Neighbours (KNN) and Decision Trees

- KNN Theory: Selection of K, Euclidean & Manhattan Distance
- Decision Trees: Classifier vs Regressor, Case Study

10: Support Vector Machines (SVM) & Clustering

- Support Vector Machine (SVM) Theory
- Case Study on SVM Implementation
- Introduction to Unsupervised Learning & Clustering
- K-Means and Hierarchical Clustering

11: Dimensionality Reduction & Hyper parameter Tuning

- Introduction to Dimensionality Reduction
- Principal Component Analysis (PCA)
- Hyper parameter Optimization Strategies: Grid Search, Random Search
- Using Scikit-Learn for Hyper parameter Tuning

12: Model Parameter and Hyperparameter Tuning

- Concepts of parameters and hyperparameters
- Manual parameter tuning methods
- Hyperparameter optimisation strategies
- Using libraries such as scikit-learn to adjust parameters
- Implementing Grid Search and Random Search
- Practical examples of hyperparameter optimisation in various algorithms

13: Handling Overfitting and Underfitting

- Definition and consequences of overfitting and underfitting
- Regularisation strategies: L1 (Lasso), L2 (Ridge)
- Enhancing accuracy through feature selection
- Implementing regularisation techniques in models
- Using libraries for feature selection
- Practical applications to mitigate overfitting and underfitting

14: Prescriptive Analytics and Optimization

- Analysing, structuring, and representing engineering problems using mathematical models
- Understanding various optimization tools to solve the problems
- Calculating feasible, relevant, and optimal solutions for these models using computational tools (GUROBI Python)
- Analysing and interpreting results to translate them into concrete actions
- Understanding the applications and limitations of linear, nonlinear, simulation, and modelling techniques in solving optimisation problems

Capstone Project: Practical Project